**BUAN 6337 Homework 4\_Group 12**

This Homework has 3 multiple part questions. You are required to use SAS for answering the

questions. Your submission on eLearning must include a pdf/word report which has followed the

sample report instructions. You should also upload your SAS code.

The SAS dataset HeinzHunts has data on grocery store purchases of Hunts and Heinz ketchup.

Each observation corresponds to one purchase occasion (of one of these brands) and consists of

the following variables:

1. Heinz : =1 if Heinz was purchased, =0 if Hunts was purchased

2. PriceHeinz : Price of Heinz

3. PriceHunts : Price of Hunts

4. DisplHeinz : = 1 if Heinz had a store display, =0 if Heinz did not have a store display

5. DisplHunts : = 1 if Hunts had a store display, =0 if Hunts did not have a store display

6. FeatureHeinz : = 1 if Heinz had a store feature, =0 if Heinz did not have a store feature

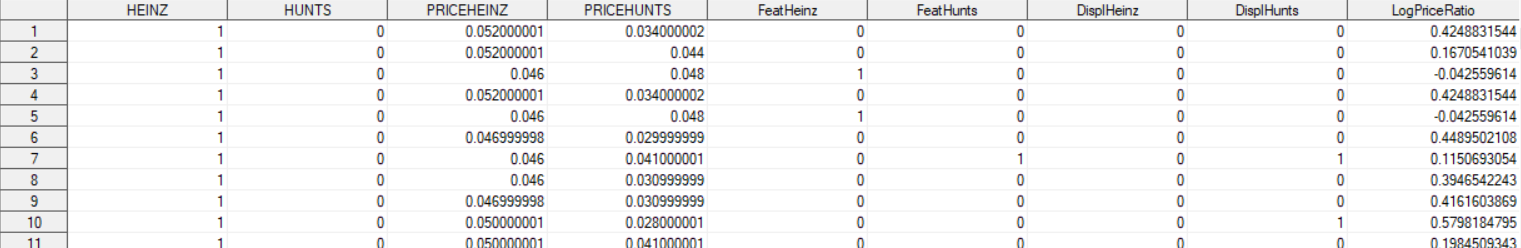
7. FeatureHunts : = 1 if Hunts had a store feature, =0 if Hunts did not have a store feature

**Question 1**

Apply the following steps and provide a screenshot of the output in your report.

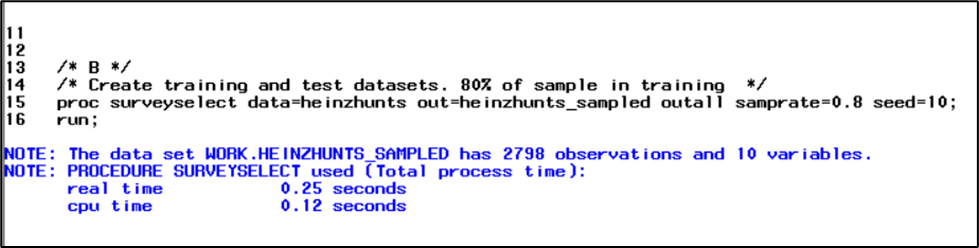
1. Create a variable LogPriceRatio = log (PriceHeinz/PriceHunts).

**Variable LogPriceRatio Added to Heinzhunts Dataset:**

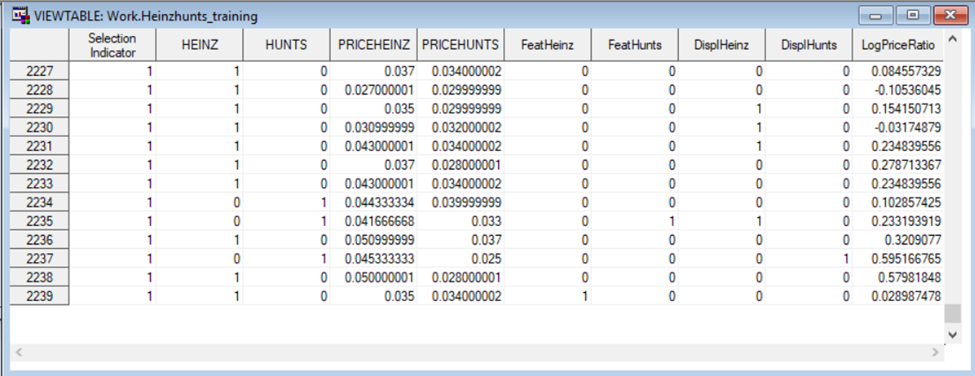


1. Randomly select 80% of the data set as the training sample, remaining 20% as test sample. (set seed=10)

**Log Output:**

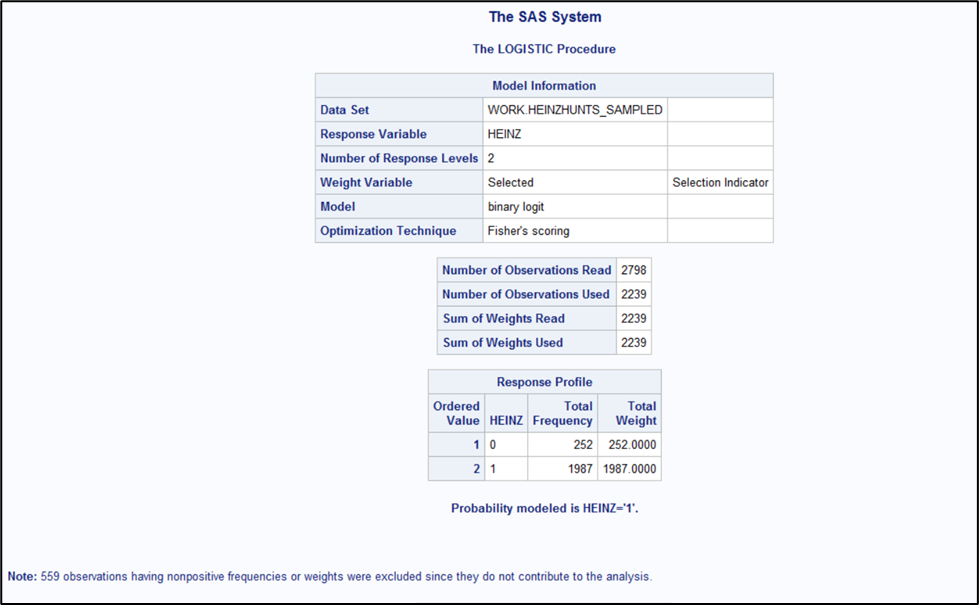


**Test Dataset Showing 80% of the total observations:**

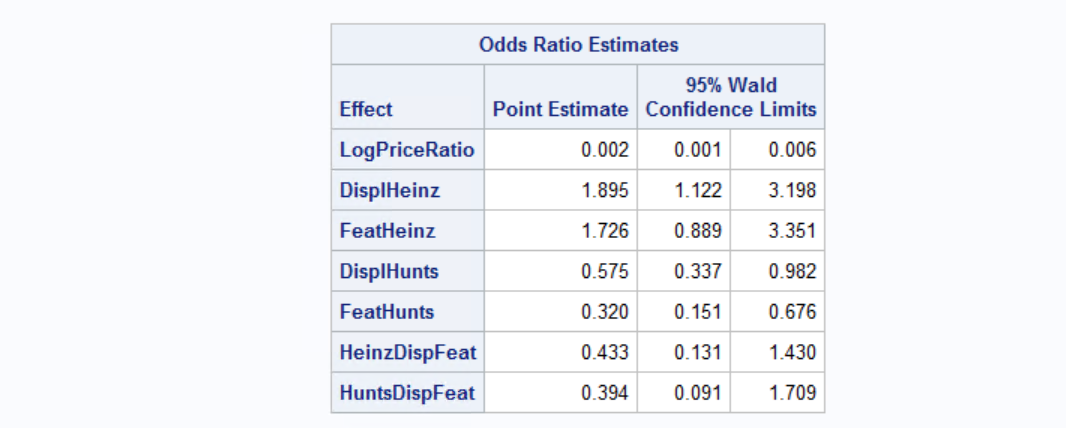


1. Estimate a logit probability model for the probability that Heinz is purchased – using LogPriceRatio, DisplHeinz, FeatureHeinz, DisplHunts, FeatureHunts as the explanatory variables. Include interaction terms between display and feature for a particular brand (e.g., DisplHeinz \* FeatureHeinz).

**Logit Model with Interaction Terms:**







1. Interpret the results. What promotional methods (feature / display) are effective for Hunts? For Heinz? How would you interpret the results for the interaction effects?

The log of odds that Heinz is purchased increases by 0.6390 and 0.5460, when Heinz makes use of a store display (vs no display for Heinz) and store feature (vs no feature for Heinz) respectively. Both are statistically significant at alpha of 0.1. Having either a display or feature for Heinz does seem to boost the likelihood of Heinz sales. Overall, Display is most effective for Heinz with an increase of 0.6390 versus an increase of 0.5460 when using a feature.

Whereas log of odds that Heinz is purchased when Hunts makes use of a store display (vs no display for Hunts) and store feature (vs no feature for Hunts) decreases by 0.5529 and 1.1403 respectively (Statistically significant at alpha of 0.05). Here the negative effect of Hunts promotions on Heinz is as expected. Overall, Hunts receives more benefit from the feature with –1.1403 compared to –0.5529 when using a display. The negative number detracting from likelihood of the customer purchasing Heinz which is in Hunts favor.

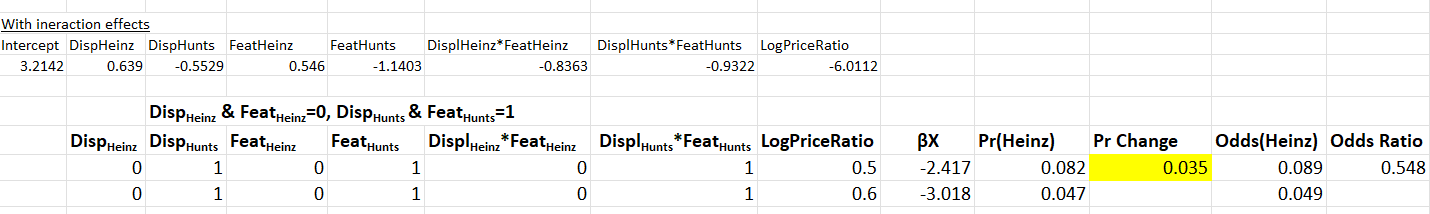
From interaction terms we infer that using both promotions at the same time does not provide statistically significant results for both brands. Here, the interaction effects for Heinz and Hunts both detract from the probability of a customer purchasing Heinz (-0.8363 for Heinz - Overuse of promotions might create a negative viewpoint in customers' mind that products sold through excessive sales promotions might either be of inadequate quality or that they are being deceived and –0.9322 for Hunts, which would be expected).

**Question 2**

Based on the estimated model in Question 1, and using the logit probability formula, calculate the change in predicted probability that Heinz is purchased if LogPriceRatio changes from 0.5 to 0.6 and Heinz does not use a feature or display, while Hunts uses a feature and a display.

The change in predicated probability that Heinz is purchased when log price ratio changes from 0.5 to 0.6 is 0.035.

**Excel Calculations for Logit Probability Model:**



**Question 3**

The above estimated model is to be used for targeting customers for Hunts coupons to build loyalty for the brand. Coupons are to be sent to customers who are likely to buy Hunts, and not to customers who are likely to buy Heinz. Therefore, the coupons should be sent to customers whose predicted probability of buying Heinz is below a certain threshold level that needs to be determined based on the costs of misclassifications (incorrectly sending / not sending a coupon)

The following information about the costs of incorrect classification is available: The cost of incorrectly sending a coupon to a customer who would have bought Heinz is $1 per customer, and the cost of incorrectly failing to send a coupon to a customer who would have bought Hunts is $0.25 per customer.

Based on these costs, what is the optimal threshold probability level that should be used with the estimated model to decide which consumers should receive coupons?

The probability threshold of 0.052318 has the lowest cost of $13.50.

**Image of Lowest Least Total Cost Variable and Probability Level:**

